

Міністерство освіти і науки України

Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"

Коляда Костянтин Вячеславович



УДК 004.934

**МЕТОДИ І ЗАСОБИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ
ВІДНОВЛЕННЯ ДАНИХ,
ВТРАЧЕНИХ ПРИ ЇХ ВІДДАЛЕНОМУ ЗБЕРІГАННІ
ТА ПЕРЕДАЧІ В МЕРЕЖАХ**

Спеціальність 05.13.05 – Комп'ютерні системи та компоненти

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2021

Дисертацією є рукопис.

Робота виконана на кафедрі системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського".

Науковий керівник: доктор технічних наук, професор
Романкевич Олексій Михайлович
КПІ ім. Ігоря Сікорського, професор кафедри
системного програмування і спеціалізованих
комп'ютерних систем

Офіційні опоненти: доктор технічних наук, професор
Бузовський Олег Володимирович
Інститут комп'ютерних технологій Університету
"Україна", професор кафедри комп'ютерної
інженерії

кандидат технічних наук, науковий співробітник
Муха Артем Андрійович
Інститут проблем математичних машин і систем
НАН України, науковий співробітник відділу
№120 "Інтегровані автоматизовані системи
спеціального призначення"

Захист відбудеться " " квітня 2021 р. о 16-30 годині на засіданні спеціалізованої ради Д 26.002.02 у КПІ ім. Ігоря Сікорського (м. Київ, проспект Перемоги 37, корп. 18, ауд. 516)

З дисертацією можна ознайомитися в бібліотеці Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського".

Відгуки на автореферат у двох примірниках, завірені печаткою установи, просимо надсилати на адресу: 03056, м. Київ, проспект Перемоги 37, вченому секретарю КПІ ім. Ігоря Сікорського.

Автореферат розісланий " " березня 2021 р.

Вчений секретар
спеціалізованої вченої ради Д 26.002.02
кандидат технічних наук, доцент

Орлова М.М.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Динамічний прогрес технологій передачі даних та глобальних мереж суттєво змінює організацію комп'ютерної обробки інформації, надаючи потужний імпульс розвитку розподілених систем обробки та зберігання даних. На сьогоднішній день масового застосування набули технології розподіленого зберігання даних користувачів на віддалених сховищах.

Важлива перевага цих технологій полягає в більшій надійності в порівнянні із зберіганням даних користувачів на їх власних накопичувачах. Інформація користувачів при їх віддаленому зберіганні потенційно може бути втрачена в результаті цілеспрямованих хакерських чи вірусних атак, помилок в роботі програмного забезпечення, відмов технічних засобів, катаклізмів природного та техногенного характеру.

Основним фактором високої надійності віддаленого зберігання інформації виступає рознесення її по різних, географічно віддалених сховищах, що локалізує втрати даних, які можуть бути відновлені за рахунок резервування. Тобто механізми резервування та відновлення втрачених даних є складовою частиною технології багаторівневого забезпечення надійності віддаленого зберігання даних користувачів.

Відповідно, динамічний розвиток систем віддаленого зберігання даних користувачів диктує необхідність адекватного вдосконалення методів та засобів резервування та відновлення втрачених даних.

Іншим важливим наслідком швидкого прогресу технології Інтернет стало розширення її використання в якості засобу обміну даними в комп'ютерних системах моніторингу стану та управління об'єктами реального світу. В якості таких об'єктів виступають побутові прилади в рамках популярної нині концепції "розумних речей", віддалені технологічні процеси, засоби відеоспостереження тощо. Для таких застосувань Інтернету критичними є надійність та своєчасність доставки пакетів. Один з найбільш дієвих способів забезпечення надійності та своєчасності доставки пакетів даних в системах реального часу полягає в застосуванні резервних пакетів, які посилаються разом з інформаційними і використовуються для відновлення втрачених або затриманих при передачі понад критичний час даних.

Таким чином, наукова задача підвищення ефективності резервування та відновлення втрачених блоків даних при їх розподіленому зберіганні на віддалених носіях чи передачі глобальною мережею в системах комп'ютерного управління є актуальною для сучасного етапу розвитку інформаційних і комп'ютерних технологій.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконувалось на кафедрі системного програмування і спеціалізованих комп'ютерних систем КПП імені Ігоря Сікорського в рамках держбюджетної теми "Методи, моделі, структури та компоненти спеціалізованих комп'ютерних систем моніторингу об'єктів критичного

застосування" (номер держреєстрації 0117U004280) згідно з науковим напрямком "Дослідження та розробка нових вискоефективних архітектур комп'ютерних систем і мереж загального і спеціального призначення, топологічної організації розподілених систем та комунікаційних технологій в них".

Мета і завдання дослідження. *Мета* роботи полягає в підвищенні ефективності відновлення даних, втрачених при їх віддаленому зберіганні або при передачі глобальною мережею.

Основні задачі дослідження у відповідності до поставленої мети полягають у наступному.

1. Аналіз особливостей сучасного та перспективного стану технологій розподіленого віддаленого зберігання інформації і транспортування пакетів даних глобальною мережею, обґрунтування критеріїв ефективності резервування даних в віддалених сховищах та при передачі мережею. Критичний огляд, з позицій визначених критеріїв, існуючих методів резервування та відновлення втрачених даних, виявлення можливостей підвищення їх ефективності, а також визначення напрямків досліджень з реалізації цих можливостей.
2. Розробка методу прискорення процесу реконструювання втрачених інформаційних блоків користувачів шляхом зменшення кількості потрібних для цього невтрачених блоків, що дозволяє знизити витрати часу на їх передачу мережею та кількість обчислювальних операцій по відновленню втрачених даних.
3. Розробка методу відновлення даних за наявності обмежень на кількість блоків, що зберігаються на одному віддаленому сховищі, який дозволяє реконструювати всю інформацію користувача в ситуації втрати доступу до одного зі сховищ, а також обмежену кількість блоків даних, що зберігаються на різних сховищах, і який забезпечує зменшення надлишковості резервування в порівнянні з відомими методами.
4. Дослідження можливості підвищення ефективності оперативного реконструювання втрачених чи затриманих понад критичний час пакетів даних в глобальних мережах, шляхом урахування статистичних характеристик їх втрат та часу передачі, а також розробка на цій основі методу відновлення втрачених пакетів даних з побудовою матриці формування резервних пакетів шляхом пріоритетного вибору її базових компонентів.
5. Розробка технології відновлення втрачених блоків даних за розробленими методами, здатної забезпечити високу швидкодію процесу реконструювання даних за рахунок вибору оптимального, в сенсі мінімуму кількості потрібних для цього операцій, варіанту відновлення, а також використання передобчислень, яке дозволяє виключити з процесу обчислення, пов'язані з розв'язанням систем рівнянь.
6. Розробка алгоритмів та програмних засобів для синтезу матриць формування резервних блоків, а також засобів для побудови таблиць

специфікацій відновлення втрачених блоків. Експериментальне дослідження ефективності запропонованих методів резервування та відновлення даних.

Об'єкт дослідження – процеси резервування інформації користувачів при довготривалому розподіленому зберіганні на віддалених сховищах чи при її передачі в глобальних мережах, а також процеси відновлення втрачених або тимчасово недоступних блоків даних чи втрачених або затриманих пакетів при їх передачі в мережах.

Предмет дослідження – методи і засоби формування резервних блоків, а також способи їх використання для відновлення втраченої інформації.

Методи дослідження базуються на теорії ймовірності та математичної статистики, теорії булевих функцій та комбінаторики, теорії завадостійкого кодування, теорії організації обчислювальних процесів, методах організації резервування, а також на використанні методів моделювання.

Наукова новизна одержаних результатів полягає в наступному.

1. Вперше запропоновано метод прискорення відновлення втраченої при віддаленому зберіганні інформації, відмінність якого полягає в формуванні резервних блоків на основі рекурентної функції Галуа з обмеженою кількістю одиничних компонент та її циклічних зсувів, що дозволяє зменшити число невтрачених блоків, потрібних для реконструкції, і тим самим знизити витрати часу на їх транспортування мережею та виконання обчислювальних операцій над ними.
2. Вперше запропоновано метод відновлення інформації при її розподіленому зберіганні не більш ніж по r блоків на кожному з віддалених сховищ, який відрізняється тим, що формування резервних блоків здійснюється на основі спеціальним чином синтезованої фрагментарно-ортогональної матриці, завдяки чому гарантується реконструкція даних при втраті доступу до одного сховища, а також трьох блоків, які зберігаються на різних сховищах, що забезпечує зменшення надлишковості резервування в порівнянні з відомими методами.
3. Вдосконалено метод відновлення втрачених чи затриманих при передачі в мережі інформаційних пакетів за рахунок використання системи пріоритетів при виборі компонентів матриці формування обмеженої кількості резервних пакетів з урахуванням статистичних характеристик затримок транспортування пакетів, що дозволяє підвищити ймовірність відновлення втраченої чи затриманої інформації і тим самим прискорити передачу даних в глобальних мережах.
4. Вперше запропоновано спосіб відновлення втрачених інформаційних блоків на основі попередньо створених таблиць специфікацій, який відрізняється використанням хеш-адресації без колізій, а також способом відновлення за критерієм мінімуму кількості невтрачених блоків, що використовуються для реконструювання даних, що дозволяє на 10-15% прискорити процес відновлення в порівнянні з відомими технологіями.

Практичне значення одержаних результатів роботи визначається тим, що їх використання дозволяє більш ефективно реалізувати технологію забезпечення надійності зберігання інформації широкого кола користувачів на віддалених сховищах. Розроблені в роботі методи дозволяють адаптувати технологію резервування та відновлення постійно чи тимчасово втрачених даних до конкретних вимог користувача та особливостей організації віддаленого зберігання інформації. Зокрема, відповідно до вимог користувача, можуть налаштовуватися часові характеристики процесу відновлення втрачених даних.

Використання результатів роботи при резервуванні та відновленні втрачених чи затриманих інформаційних пакетів в глобальних мережах дозволяє прискорити доставку даних. Це має велике практичне значення для систем моніторингу стану віддалених об'єктів та комп'ютерного управління ними в реальному часі, які використовують глобальні мережі в якості середовища передачі даних.

Особистий внесок здобувача полягає в теоретичному обґрунтуванні одержаних результатів, експериментальній їх перевірці та дослідженні, а також в створенні програмних продуктів для практичного використання одержаних результатів.

Всі результати, що наведені в дисертації, отримані автором самостійно. У роботах, що написані в співавторстві, автору належать: [1] – метод відновлення інформації при втраті доступу до одного сховища, а також трьох блоків, які зберігаються на різних сховищах; [2] – спосіб прискореного відновлення втрачених інформаційних блоків, [3] – вдосконалення методу відновлення затриманих при передачі інформаційних пакетів даних, за рахунок використання системи пріоритетів при виборі компонентів матриці формування резервних пакетів; [4] – обґрунтування вибору критеріїв ефективності резервування даних та порівняльний аналіз, з позиції обраних критеріїв, відновлення даних з використанням резервування та корегуючих кодів; [5] – спосіб використання попередньо створених таблиць специфікацій для ефективного відновлення втраченої інформації; [6] – порівняльний аналіз відновлюючих та корегуючих кодів; [7] – підхід до відновлення всіх блоків даних користувача, що зберігаються на одному сховищі; [8] – метод прискорення відновлення втраченої при віддаленому зберіганні інформації за рахунок розрідження матриці формування резервних блоків даних; [9] – підхід до відновлення втрачених пачок пакетів даних в мережах.

Дисертаційна робота виконана на кафедрі системного програмування і спеціалізованих комп'ютерних систем КПП ім. Ігоря Сікорського. Науковий керівник д.т.н., професор Романкевич Олексій Михайлович.

Апробація результатів дисертації. Основні результати дисертації доповідались та обговорювались на 6 науково-технічних конференціях, в тому числі двох міжнародних:

1. International Conference of World Scientific and Engineering Academy and Society - Mathematics and Computers in Sciences and Industry. Aug. 24-26, 2019, Corfu Island, Greece.

2. III-й Міжнародній науково-практичній конференції "Проблеми інформатики та комп'ютерної техніки", 27-31 травня 2013р., Чернівці.
3. III-й Українській конференції з автоматичного керування "Автоматика-96". 12-16 травня 1996 р., м. Севастополь.
4. Всеукраїнській науково-практичній конференції "Передача и обработка данных в системах управления и сетях". 7-9 лютого 1989 р., м. Київ.
5. Науково-технічній конференції "Актуальные проблемы в области радиоэлектроники, автоматики, вычислительной техники, энергетики, машиностроения и промышленных технологий", 24-26 жовтня 1988 р., м. Київ.
6. VI Координаційній нараді "Развитие методов проектирования и изготовления интегральных запоминающих устройств", 2-3 лютого 1988 р., м. Київ.

Публікації. Основні положення дисертаційної роботи опубліковані в 9 наукових працях, серед яких 7 статей у наукових фахових виданнях (в тому числі одна стаття, яка реферується міжнародними наукометричними базами даних SCOPUS, ISI Thomson's Scientific and Technical Proceedings, ISTP/ISI Proceedings, EI's Engineering Information Index, EI Compendex, DBLP, ACM, Google Scholar та IEEE Xplore), 6 статей, що входять до наукометричної бази даних РІНЦ та тези доповідей на конференціях.

Структура та об'єм роботи. Дисертаційна робота складається з вступу, чотирьох розділів, висновків та додатків. Загальний обсяг роботи складає 141 сторінку, робота містить 4 рисунки, 11 таблиць та список використаної літератури із 78 найменувань, 2 додатки.

ОСНОВНИЙ ЗМІСТ

У вступі обґрунтована актуальність проблеми підвищення ефективності резервування і відновлення даних в системах розподіленого віддаленого зберігання інформації, які динамічно розвиваються в рамках перспективних "хмарних" технологій. Показано, що аналогічні задачі резервування та відновлення пакетів даних постають в глобальних мережах, які з появою дешевих радіомодемів почали широко використовуватися в якості середовища обміну даними в системах віддаленого управління, що працюють в реальному часі. Для цих застосувань, передача резервних пакетів та швидке відновлення затриманих або втрачених інформаційних пакетів виступає чинником підвищення оперативності доставки даних.

Для забезпечення надійного віддаленого зберігання даних користувачів в розподілених системах та оперативного доступу до них потрібні спеціальні механізми резервування та відновлення даних, доступ до яких постійно або тимчасово втрачено.

Формуються мета та задачі дослідження, визначаються наукова новизна та практичне значення одержаних результатів.

У першому розділі дисертації проаналізовано особливості сучасного та перспективного стану технологій розподіленого віддаленого зберігання інформації і транспортування пакетів даних глобальною мережею.

Інформація, що зберігається на носіях, може бути втрачена в результаті дії вірусів, направлених хакерських атак, виходу з ладу технічного обладнання, помилок та збоїв в роботі програмного забезпечення, катаклізмів природного, техногенного та військового характеру.

Зростаюче розповсюдження віддаленого зберігання даних в рамках хмарних технологій зумовлене меншою ймовірністю втрати інформації, наявністю значних за обсягом ресурсів пам'яті, простотою обміну та розповсюдження даних. Підвищення надійності зберігання інформації зумовлене, головним чином, за рахунок розподілення даних різними сховищами: наведені вище причини втрати не можуть одночасно реалізуватися на географічно рознесених вузлах і саме це обмежує об'єм втраченої інформації, яка може бути відновлена за рахунок резервування.

В рамках окремого вузла організовано розподілення даних по носіях, доступ користувачів до даних, їх захист та резервування.

Очевидно, що резервування на рівні одного вузла практично не дозволяє зменшити ймовірність втрат інформації, викликаних катаклізмами, хакерськими атаками та збоями в роботі програмного забезпечення.

Потужний прогрес Інтернету надає можливість здійснювати резервування на рівні користувача. При цьому повною мірою можна задіяти можливості рознесеності зберігання інформації для локалізації її можливих втрат. Тільки цей чинник здатен реально забезпечити захист даних від втрат внаслідок катаклізмів. Проте рознесення зберігання даних негативно впливає на швидкість доступу до даних – їх потрібно зібрати з різних сховищ.

Тому для підвищення надійності зберігання даних користувачів з урахуванням їх специфічних вимог, резервування має здійснюватися на рівні, вищому ніж рівень окремого вузла.

Проблема надійності передачі даних посідає чільне місце в мережевих технологіях від самого початку їх розвитку. В сучасних мережах використовується дві базові технології: виправлення обмеженої кількості помилок з використанням корегуючих кодів, які локалізують і виправляють спотворені символи пакету, та повторна передача пакету.

Виявлення та виправлення помилок може здійснюватися при кожній передачі пакету між вузлами комутації на всьому шляху його доставки. Така організація дозволяє виключити накопичення помилок, але при цьому суттєво уповільнює процес доставки пакету.

Сьогодні, коли глобальні мережі використовуються в якості середовища обміну даних системами управління віддаленими об'єктами в реальному часі, існуючі технології забезпечення надійності, які зумовлюють затримки доставки пакетів, не задовольняють вимогам оперативності управління.

Оскільки основним чинником затримки доставки пакетів виступають повторні передачі, то більш ефективним рішенням є посилка разом з інформаційними певної кількості резервних пакетів для швидкого відновлення тих інформаційних пакетів, які запізнились понад критичну часову межу.

Проведений аналіз показав, що в сучасних умовах в якості критеріїв ефективності систем резервування виступають:

- кількість інформаційних блоків, які можуть бути відновлені;

- кількість резервних блоків;
- час відновлення даних, який визначається головною мірою часом транспортування блоків даних, інформаційних та резервних, що використовуються при реконструкції втрачених даних.

Широкого поширення набули технології дублювання та резервування даних, які реалізовані в сімействі систем RAID (Redundant Arrays of Independent Disks). RAID-5 система забезпечує відновлення даних при втраті одного носія. Система RAID-6 відновлює дані при їх втраті з двох носіїв.

Недоліки цих систем полягають в тому, що вони не враховують специфічних вимог до надійності доступу до даних та оперативності доступу до них кожного користувача.

Для відновлення даних при їх віддаленому зберіганні переважно використовуються MDS (Maximum Distance Separable) коди. Для відновлення двох чи більшої кількості втрачених блоків можуть бути використані різні різновиди MDS-кодів: коди Ріда Соломона, матричні коди типу EVENODD, RDP, Х-коди, В-коди, С-коди, Liberation-коди.

Проведений аналіз показав, що недоліки цих кодів полягають в тому, що вони не враховують специфіки реального віддаленого рознесеного зберігання інформації, а також не забезпечують можливість гнучкого пристосування до зміни критеріїв ефективності для кожного конкретного застосування.

Практика використання віддаленого зберігання даних постійно диктує нові вимоги. Зокрема, в останні роки динамічно зростає кількість користувачів, які використовують хмарні сервіси для зберігання інформації. Для багатьох застосувань критичним є швидкість доступу до віддаленої інформації. Технологія відновлення даних стала активно використовуватися для швидкого реконструювання даних, які доставляються з затримками.

Вказані фактори потребують розробки адекватних сучасній ситуації підходів до вдосконалення технологій резервування та відновлення даних.

В другому розділі обґрунтовано та розроблено підходи до підвищення ефективності відновлення даних, втрачених на віддалених сховищах.

У рамках роботи досліджується наступна відома модель збереження інформації на віддалених носіях. Користувач зберігає на віддалених серверах n інформаційних блоків B_1, B_2, \dots, B_n . Для їх відновлення, в разі втрати доступу до них, формується m резервних блоків R_1, R_2, \dots, R_m у вигляді лінійних перетворень над інформаційними блоками:

$$\forall l \in \{1, 2, \dots, m\} : R_l = \bigoplus_{j=1}^n a_{l,j} \cdot B_j ,$$

де $a_{l,j} \in \{0, 1\}$ – бінарні коефіцієнти матриці A , яка визначає порядок формування резервних блоків даних:

$$A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{vmatrix} .$$

В реальних системах віддаленого зберігання інформації користувачів та системах комп'ютерного управління віддаленими об'єктами з використанням Інтернету в якості середовища обміну даних, важливу роль відіграє час відновлення втрачених інформаційних блоків чи затриманих доставкою понад критичний час пакетів даних.

Для систем віддаленого зберігання даних час відновлення складається з двох складових: часу доставки резервних та інформаційних блоків, потрібних для процесу відновлення, та часу виконання обчислень по реконструюванню втрачених блоків. Для систем комп'ютерного управління з обміном даними через Інтернет, час відновлення втрачених чи затриманих пакетів визначається лише часом реалізації обчислень.

Показано, що середній час T_m відновлення блоку даних при зберіганні інформації в хмарах визначається формулою:

$$T_m = \varepsilon \cdot t_T + (\varepsilon - 1) \cdot t_{XOR}, \quad (1)$$

де ε – середня кількість інформаційних блоків, з яких формується один резервний блок, t_T – час доставки блоку, що віддалено зберігається, t_{XOR} – час логічного додавання двох блоків; на практиці $t_{XOR} \ll t_T$.

З формули (1) випливає, що найбільш дієвий шлях прискорення відновлення втрачених блоків чи недоставлених до встановленого часу пакетів, полягає в зменшенні кількості ε пакетів, з яких формується кожен з резервних блоків (пакетів). Виходячи з цього, теоретично найменший час відновлення втрачених блоків досягається при використанні дублювання інформаційних блоків, тобто при $\varepsilon=1$. Проте дублювання має наслідком значну надлишковість: так для гарантування відновлення h втрачених блоків від загальної кількості n інформаційних блоків потрібна кількість m резервних блоків становить $m=h \cdot n$.

Для того, щоб забезпечити відновлення h інформаційних блоків, передбачається формування резервних блоків з використанням систем квазістаціонарних рекурентних функцій Галуа. Ці функції базуються на ортогональних системах функцій базису Уолша. Системи дискретних функцій Уолша являють собою системи ортонормованих прямокутних бінарних функцій. Існують різні способи впорядкування систем такого класу. При цьому з базису Уолша системи трансформуються в базиси Галуа чи Адамара. В роботі передбачається використання систем в базисі Галуа.

В рамках цього відомого підходу запропонована модифікована процедура побудови матриці формування резервних блоків. Для задачі гарантованого відновлення інформаційних блоків при втраті не більше h із них потрібні прямокутні матриці, що складаються з n стовпців та m рядків.

Сутність використання систем квазістаціонарних рекурентних функцій Галуа для формування MDS кодів полягає в тому, що спочатку формується бітова послідовність $S(n)$, яка являє собою перший рядок матриці A . Наступні μ рядків матриці формуються циклічним зсувом послідовності $S(n)$. Стартовий код $S_0 = \{s_1, s_2, \dots, s_n\}$ вибирається таким чином, щоб сума за модулем два будь-якої пари кодів, утворених циклічним зсувом коду S не співпадала ні з одним циклічно зсунутим кодом S_0 . Іншими словами, якщо

позначити через U множину n -бітових кодів, утворених циклічним зсувом стартового коду S_0 : $U = \{S_0, S_1, \dots, S_{n-1}\}$, то для будь якої пари S_i, S_j кодів, що належать U : $S_i, S_j \in U, \forall i, j \in \{0, 1, \dots, n-1\}$ їх сума за модулем два не належить множині U : $S_i \oplus S_j \notin U$.

$$\forall S_i, S_j \in U, i, j \in \{0, 1, \dots, n-1\}: S_i \oplus S_j \notin U$$

Якщо бінарний код S_0 містить в собі фіксовану кількість ε одиниць, то потрібно підібрати стартовий код S_0 таким чином, щоб при циклічному зсуві співпадала мінімальна кількість одиниць.

Для підвищення ефективності цієї технології пропонується використувати вектор $E = \{e_1, e_2, \dots, e_\varepsilon\}$ – вектор відстаней між позиціями одиниць в бінарному коді S_0 . Це означає, що e_1 – різниця позицій другої та першої одиниць в коді S , e_2 – різниця номерів позицій третьої та другої одиниць в коді S ,... e_ε – циклічна різниця між позиціями останньої та першої одиниць в коді S ; цілком очевидно, що $e_1 + e_2 + \dots + e_\varepsilon = n$. Кількість ε одиниць в коді S_0 визначається як арифметична сума бітів цього коду: $\varepsilon = s_1 + s_2 + \dots + s_n$.

Основна перевага використання вектора E полягає в тому, що при зсуві відносна відстань між одиницями циклічно зсувного коду S_0 не змінюється. Може змінюватися тільки відносний порядок одиниць.

Для прискорення процесу відновлення втрачених даних за рахунок зменшення кількості інформаційних та резервних блоків, які використовуються для реконструкції втрачених блоків, пропонується модифікована процедура формування резервних блоків. Ця процедура передбачає побудову матриці A таким чином, щоб кожен її рядок містив ε одиниць, де ε – кількість блоків, які потрібно транспортувати мережею для відновлення інформаційного блоку. При цьому збільшується кількість m потрібних для відновлення резервних блоків.

Показано, що при малих значеннях $\varepsilon \ll n/2$, виникає ситуація коли $m > n$. В цьому випадку пропонується використовувати додаткову бітову послідовність $S'(n)$, яка утворює $(n+1)$ -й рядок матриці A , а всі інші рядки після нього утворюються циклічним зсувом послідовності $S'(n)$. Виявлені умови вибору послідовності $S'(n)$ з урахуванням першої послідовності $S'(n)$.

При заданому значенні ε та h -кількості блоків, що мають відновлюватися при втраті, число m резервних блоків визначається з формули:

$$m = \max\left\{\frac{n \cdot h}{\varepsilon}, (h-2) \cdot \lceil \log_2(n+1) \rceil + 2\right\}. \quad (2)$$

Наприклад, при $n=6$ та $h=3$ можна розглянути та порівняти два варіанти побудови матриці A формування резервних блоків: при $\varepsilon = 3$ і $\varepsilon=2$. В першому варіанті згідно формули (2) $m=6$, а в другому $m=9$. В якості стартового коду S_1 першої матриці можна вибрати $S_1 = \{1, 1, 0, 1, 0, 0\}$. Для побудови матриці A при $m=9$ згідно з запропонованим методом використовується дві послідовності S_2 та S_2' , які формуються векторами $E_2 = \{2, 4\}$ та $E_2' = \{1, 5\}$. Відповідно, стартові коди дорівнюють: $S_2 = \{1, 0, 1, 0, 0, 0\}$, і $S_2' = \{1, 1, 0, 0, 0, 0\}$.

Матриці A_1 (при $\varepsilon = 3$) та A_2 (при $\varepsilon=2$) формування резервних блоків мають наступний вигляд:

$$A_1 = \begin{vmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{vmatrix} \quad A_2 = \begin{vmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix}$$

В разі втрати другого, третього та четвертого блоків, відповідні стовпці матриць A_1 та A_2 включають відповідно ортогональні підматриці Θ_1 і Θ_2 :

$$\Theta_1 = \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} \quad \Theta_2 = \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix}$$

Відновлення втрачених блоків відбувається для матриці A_1 у вигляді:

$$B_2 = R_3 \oplus B_5 \oplus B_6; \quad B_3 = R_2 \oplus B_1 \oplus B_6; \quad B_4 = R_4 \oplus B_1 \oplus B_5$$

Відновлення втрачених блоків для матриці A_2 здійснюється в такій послідовності:

$$B_4 = R_2 \oplus B_6; \quad B_3 = R_7 \oplus B_4; \quad B_2 = R_3 \oplus B_4$$

В рамках наведеного прикладу при $\varepsilon=3$ для відновлення трьох блоків потрібно здійснити доставку 6-ти блоків та виконати 6 операцій логічного додавання блоків. При $\varepsilon=2$ для відновлення цих же блоків потрібно доставити 4 блоки та виконати три операції логічного додавання.

Таким чином, час відновлення в другому випадку більш ніж в 1.5 рази менший, ніж при $\varepsilon=3$, за рахунок збільшення в 1.5 рази числа резервних блоків.

В загальному випадку при використанні запропонованого методу коефіцієнт ξ прискорення відновлення визначається формулою:

$$\xi = \frac{n \cdot h}{\varepsilon \cdot (h-2) \cdot \lceil \log_2(n+1) \rceil + 2}. \quad (3)$$

Як зазначалося вище, вагомим чинником забезпечення надійності віддаленого зберігання даних є їх рознесення по різних сховищах. Це обмежує одномоментну втрату інформації об'ємом, що зберігається на одному сховищі. В роботі запропоновано метод, який дозволяє гарантовано відновлювати всі блоки даних користувача, що зберігаються на одному сховищі за умови, що їх кількість обмежена значенням r .

Крім гарантування можливості реконструкції всіх r блоків даних при втраті доступу до одного сховища, де вони зберігаються, метод гарантує відновлення до трьох блоків, які зберігаються на різних сховищах. За рахунок орієнтації на вказаному типі втрат даних метод дозволяє підвищити ефективність шляхом зниження надлишковості резервування, тобто він для

вирішення вказаної вище задачі відновлення використовує меншу кількість резервних блоків в порівнянні з універсальними.

Розроблений метод передбачає, що стовпці матриці A розділяються на $g = \lceil n/r \rceil$ фрагментів так, що в перших $g-1$ фрагментах міститься по r стовпців, а g -й фрагмент містить $n-r \cdot (g-1)$ стовпців.

Побудову матриці A пропонується здійснювати наступним чином.

1. Встановити індекс j в нуль : $j=0$.
2. $(i+j) \bmod r$ -тий рядок i -того фрагменту матриці A , $i \in \{1, 2, \dots, g\}$, заповнюється $(r-j)$ -розрядним кодом, що складається з одиниць на $(r+j-1-i) \bmod r$ -й позиції та циклічно наступних за нею в рамках r -бітового фрагменту $r-1-j$ нулів.
3. Здійснюється інкремент індексу j : $j=j+1$. Якщо $j < r$, то виконується перехід на повторне виконання п.2.
4. Невизначені елементи перших $\lceil \log_2 (n+1) \rceil$ рядків матриці A заповнюються таким чином, щоб для кожного стовпця вони утворювали унікальний номер, починаючи з одиниці.
5. Наступні рядки матриці A заповнюються таким чином, щоб кількість одиниць в кожному зі стовпців матриці A була не меншою трьох.

Наприклад, для $r=4$ і $n=15$ матриця A формування резервних блоків, побудована за розробленим методом, має наступний вигляд:

$$A = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

При втраті всіх блоків B_5, B_6, \dots, B_8 , які зберігаються на 2-му сховищі, відповідні стовпці матриці A містять в собі ортогональну підматрицю Θ :

$$\Theta = \begin{vmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{vmatrix}$$

Відповідно, втрачені блоки відновлюються наступним чином:

$$B_7 = R_2 \oplus B_1 \oplus B_4 \oplus B_{11} \oplus B_{12} \oplus B_{14}$$

$$B_6 = R_5 \oplus B_4 \oplus B_1 \oplus B_2 \oplus B_3 \oplus B_9 \oplus B_{10} \oplus B_{12} \oplus B_{15}$$

$$B_8 = R_3 \oplus B_7 \oplus B_2 \oplus B_4 \oplus B_{10} \oplus B_{13} \oplus B_{14} \oplus B_{15}$$

$$B_5 = R_4 \oplus B_7 \oplus B_8 \oplus B_1 \oplus B_3 \oplus B_{10} \oplus B_{11} \oplus B_{13}$$

При втраті трьох блоків B_1, B_7, B_{12} , що зберігаються на різних сховищах, відповідні стовпці матриці A містять ортогональну підматрицю Θ :

$$\Theta = \begin{vmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix}$$

Відповідно, втрачені блоки відновлюються так:

$$B_7 = R_3 \oplus B_2 \oplus B_4 \oplus B_8 \oplus B_{10} \oplus B_{13} \oplus B_{14} \oplus B_{15}$$

$$B_{12} = R_1 \oplus B_7 \oplus B_4 \oplus B_5 \oplus B_6 \oplus B_8 \oplus B_{11} \oplus B_{14}$$

$$B_1 = R_5 \oplus B_2 \oplus B_3 \oplus B_6 \oplus B_9 \oplus B_{10} \oplus B_{12} \oplus B_{15}$$

Основний ефект запропонованого методу полягає в тому, що суттєвим чином зменшується надлишковість резервування в порівнянні з відомими методами за рахунок урахування особливостей втрати доступу до даних при їх віддаленому зберіганні. Показано, що надлишковість резервування зменшується в μ раз, чисельне значення μ визначається формулою:

$$\mu = \frac{(r-2) \cdot \log_2(n+1) + 2}{r+2}. \quad (5)$$

Для наведеного прикладу число резервних блоків для відновлення 4-х, довільно локалізованих по сховищам блоків, дорівнює 10, а в запропонованому методі, який дозволяє відновлювати 4 блоки, що зберігаються на одному сховищі, і три, довільно локалізовані блоки, потрібно 6 резервних блоків. Відповідно, надлишковість резервування зменшується $\mu = 1.67$ раз.

Третій розділ роботи присвячено підвищенню ефективності відновлення втрачених та затриманих понад критичний час інформаційних пакетів в глобальних мережах.

Одним із дієвих резервів підвищення ефективності оперативного реконструювання втрачених чи затриманих понад критичний час пакетів в глобальних мережах є урахування статистичних характеристик їх втрат та часу передачі. Для практичного використання цього резерву розроблено метод формування резервних пакетів та відновлення з їх використанням втрачених або затриманих понад критичний час основних пакетів даних.

Метод використовує систему пріоритетів при виборі компонентів матриці, формування обмеженої кількості резервних пакетів з урахуванням статистичних характеристик, затримок транспортування пакетів, що дозволяє підвищити ймовірність відновлення втраченої чи затриманої інформації і тим самим прискорити передачу даних в глобальних мережах.

Проведені статистичні дослідження довели, що запропонований метод дозволяє на 7-10% підвищити ймовірність реконструювання втрачених чи затриманих понад критичний час пакетів в порівнянні з відомими методами.

Проведений аналіз показав, що при передачі даних в мережах з ефірними каналами типовою є ситуація втрата "пачки" суміжних пакетів потоку (BPL - Burst Packet Loss) під дією зовнішньої електромагнітної завади.

Для ефективного відновлення пакетів даних розроблено спеціальний метод їх прискореного відновлення. Метод передбачає побудову матриці А формування резервних пакетів з двох частин, перша з яких гарантує відновлення всіх пакетів "пачки", а друга – двох довільних пакетів.

За рахунок спеціалізації на найбільш типовому варіанті втрати пакетів в системах з ефірними та супутниковими каналами метод має підвищену ефективність у порівнянні з універсальними рішеннями за рахунок зменшення потрібних для відновлення резервних пакетів, та збільшення

швидкодії. Прискорення відновлення BPL досягається за рахунок того, що підматриця A формування резервних пакетів, призначених для реконструювання BPL, містить меншу кількість одиниць ніж матриця A , що містить стандартні MDS коди. Якщо стандартна матриця MDS кодів містить, в середньому, $n/2$ одиниць рядку, то рядок матриці A , побудованої за запропонованим способом містить n/g одиниць, де g -максимальна кількість пакетів в "пачці". Середня кількість N операцій логічного додавання для відновлення в запропонованому методі визначається формулою:

$$N = R_b \cdot \frac{n}{g} = R_b \cdot \frac{n}{3 \cdot \sqrt{R_b}} = \frac{n \cdot \sqrt{R_b}}{3}, \quad (6)$$

де R_b - середнє значення числа пакетів в втраченій "пачці".

При використанні універсального методу відновлення BPL з використанням стандартних MDS кодів, для відновлення R_b інформаційних пакетів потрібно виконати $N_0 = R_b \cdot n/2$ операцій логічного додавання інформаційних частин пакетів. Це означає, що використання запропонованого спеціалізованого методу для вирішення задачі відновлення BPL дозволяє прискорити процес відновлення пакетів в ζ раз, причому чисельне значення коефіцієнта ζ прискорення визначається наступною формулою:

$$\zeta = \frac{N_0}{N} = \frac{3 \cdot \sqrt{R_b}}{2}. \quad (7)$$

Значення R_b - середнього числа пакетів із втраченої "пачки", на практиці лежить в інтервалі від 3-х до 10-х. Відповідно, теоретично обчислене за формулою (7) значення коефіцієнта ζ прискорення реконструювання втрачених пакетів BPL для реальних ситуацій лежить в інтервалі від 2.6 до 4.7. Експериментальні дослідження показали, що реальне прискорення дещо менше, але близьке до отриманих теоретичним шляхом оцінок.

Запропоновані в дисертаційній роботі механізми відновлення затриманих даних, за рахунок вчасно доставлених резервних блоків, ні в якій мірі не заперечують використання протоколів TCP/IP по захисту даних від помилок та втрат, оскільки діють на іншому мережному рівні.

Існуючі механізми виправлення помилок при передачі пакету чи відновлення втрачених пакетів шляхом їх повторної передачі на транспортному рівні, в певних умовах, вносять значну затримку. Зокрема, при використанні супутникових каналів, які характеризуються високим рівнем втрат та великими значеннями RTT (Round Triple Time), затримка, зумовлена роботою протоколів транспортного рівня, може становити 100-300 мс. Запропоновані механізми відновлюють втрачені дані швидше: за 5-10 мс.

Крім того, існують ситуації коли протоколи виправлення помилок та втрат транспортного рівня використовувати не можливо. Для цих ситуацій запропонований механізм дозволяє відновлювати втрачену інформацію. Реалізація протоколів TCP/IP щодо виправлення помилок на термінальному рівні не завжди забезпечується відповідними апаратними засобами.

Четвертий розділ роботи присвячено розробці технології відновлення блоків даних при застосуванні методів, запропонованих в другому та третьому розділах дисертаційної роботи.

В технологічному плані процес відновлення втрачених інформаційних блоків полягає в розв'язанні системи лінійних рівнянь, коефіцієнти якої утворені компонентами матриці Θ . Аналіз трансформованої матриці A на предмет виявлення в ній ортогональної підматриці Θ , формування на основі матриці Θ системи лінійних рівнянь з подальшим її розв'язанням потребує певних обчислювальних і часових ресурсів, що критично для відновлення втрачених даних в реальному часі.

Кількість q варіантів відновлення h втрачених блоків даних залежить від загального числа n інформаційних блоків та числа m резервних блоків, які використовуються для реконструювання. Кількість варіантів відновлення втрачених блоків даних значно зростає, якщо кількість блоків, що реконструюються помітно менша за граничне значення.

Проведений аналіз показує, що за рахунок оптимізації вибору варіанту відновлення за критерієм мінімуму кількості блоків, які потрібно транспортувати глобальною мережею, можна прискорити процес відновлення даних, в середньому на 15%, причому ефективність такої оптимізації зростає зі збільшенням кількості блоків даних, які зберігаються віддалено.

Оскільки виділення з матриці A ортогональної підматриці Θ та її подальше розв'язання потребує певних часових ресурсів, що не завжди є прийнятним для систем, що працюють в реальному часі, пропонується табличний спосіб відновлення втрачених пакетів.

Пропонований спосіб передбачає попереднє створення таблиці специфікацій, які визначають для кожної комбінації втрачених блоків (інформаційних та резервних) перелік невтрачених блоків, потрібних для реконструювання та специфікацію, яка описує обчислювальні операції по відновленню втрачених блоків.

Відповідно, на етапі налаштування вибирається така ортогональна матриця, яка дозволяє оптимізувати процес відновлення за критеріями мінімуму блоків, що мають бути транспортованими для відновлення втрачених блоків, та обчислювальної складності операцій реконструювання.

Кількість специфікацій N_c визначається числом комбінацій втрачених блоків, серед яких є інформаційні: при втраті u блоків з загальної кількості n основних і m резервних блоків, кількість комбінацій за участю інформаційних блоків дорівнює $C_{m+n}^u - C_m^u$.

Відповідно, загальна кількість N_c специфікацій відновлення втрачених основних блоків, що містяться в таблиці, за умови що їх кількість не перевищує k , визначається такою формулою:

$$N_c = \sum_{u=1}^k \sum_{j=1}^u C_n^j \cdot C_m^{u-j} \cdot j. \quad (8)$$

Відповідно, об'єм V пам'яті в бітах, яку займає таблиця специфікацій відновлення основних блоків даних, визначається формулою:

$$V = N_c \cdot (m + n) = (m + n) \cdot \sum_{u=1}^k \sum_{j=1}^u C_n^j \cdot C_m^{u-j} \cdot j. \quad (9)$$

Дані про кількість специфікацій та потрібний для їх зберігання об'єм пам'яті для різної кількості інформаційних блоків за умови їх гарантованого відновлення при втраті не більше трьох блоків наведено в таблиці 1.

Таблиця 1

Залежність кількості специфікацій гарантованого відновлення всіх основних пакетів за умови втрати не більше трьох переданих пакетів.

Кількість n основних пакетів	Мінімальна кількість m резервних пакетів	Число N специфікацій	Об'єм V пам'яті специфікацій, байт
7	5	469	703
15	6	3165	8308
31	7	21824	$103 \cdot 10^3$
63	8	156618	$139 \cdot 10^4$
127	9	$116 \cdot 10^4$	$198 \cdot 10^5$

Аналіз даних, наведених в таблиці 1, свідчить про те, що об'єми пам'яті таблиць специфікацій лежать у межах можливості технічної реалізації навіть з використанням обчислювальних платформ відносно невеликої потужності.

На основі проведених експериментальних досліджень показано, що запропонований спосіб дозволяє на 10-15% прискорити обчислювальну реалізацію реконструювання втрачених блоків в порівнянні з відомими технологіями.

В додатках наведені лістинги розроблених програмних засобів.

ВИСНОВКИ

В дисертаційній роботі виконано теоретичне обґрунтування і одержано нове вирішення наукової задачі підвищення ефективності резервування та відновлення втрачених даних в розподілених системах зберігання інформації, а також при передачі в глобальних мережах.

Основні наукові і практичні результати полягають у наступному.

1. Виконано аналіз сучасного стану та перспектив розвитку технологій розподіленого віддаленого зберігання інформації та транспортування пакетів даних глобальною мережею, обрано критерії ефективності засобів резервування та відновлення втрачених даних в віддалених сховищах і при передачі мережею. Показано, що в сучасних умовах зростає значимість часу відновлення втрачених даних. Виявлено, що дієвим резервом підвищення ефективності резервування та відновлення даних є урахування особливостей організації їх віддаленого зберігання в реальних системах.
2. Розроблено метод прискорення відновлення втрачених блоків даних, який дозволяє, за рахунок збільшення надлишковості резервування, зменшити час доставки глобальною мережею інформаційних та резервних блоків для

реконструювання втрачених блоків, а також зменшити кількість потрібних для цього обчислювальних операцій, що забезпечує важливе для багатьох застосувань прискорення відновлення даних.

3. Теоретично обґрунтовано, розроблено та досліджено метод відновлення інформаційних блоків при їх розподіленому зберіганні не більш ніж по r на кожному з віддалених сховищ, який базується на використанні для формування резервних блоків матриці, що є ортогональною в межах суміжних r стовпців, а також будь-які три стовпці якої містять ортогональну підматрицю, що гарантує реконструкцію всіх r блоків при втраті доступу до одного сховища, на якому вони зберігаються, а також трьох блоків, які зберігаються на різних сховищах. Метод використовує меншу кількість резервних блоків, тобто має нижчий рівень надлишковості резервування в порівнянні з відомими методами, орієнтованими на відновлення будь-яких r втрачених блоків.

4. Теоретично обґрунтовано, розроблено та досліджено метод відновлення втрачених при передачі в глобальних мережах пакетів даних, який відрізняється використанням системи пріоритетів при виборі компонентів матриці формування обмеженої кількості резервних пакетів з урахуванням статистичних характеристик втрат пакетів та затримок, що дозволило на 7-10% підвищити ймовірність можливості реконструювання втрачених чи затриманих понад критичний час пакетів в порівнянні з відомими методами.

5. Запропоноване ефективне технологічне рішення обчислювальної реалізації створених методів у вигляді способу відновлення втрачених інформаційних блоків на основі попередньо створених таблиць специфікацій, який за рахунок оптимізації вибору варіанту реконструювання за критерієм мінімуму кількості блоків даних, які використовуються для цього, а також за рахунок застосування передобчислень, що разом виключають з процесу безпосереднього відновлення ті обчислення, які пов'язані з розв'язанням систем рівнянь, дозволяє на 10-15% прискорити обчислювальну реалізацію реконструювання втрачених блоків в порівнянні з відомими технологіями.

6. Розроблено програмні засоби для реалізації синтезу матриць формування резервних блоків за розробленими методами, а також алгоритми та програми для побудови таблиць специфікацій відновлення втрачених блоків. З використанням розроблених програмних засобів проведено експериментальне дослідження ефективності запропонованих методів резервування та відновлення даних, яке, в цілому, підтвердило отримані теоретично результати.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Коляда К.В. Метод відновлення даних при їх розподіленому зберіганні на віддалених сховищах / К.В. Коляда, В.О. Романкевич, М.М. Орлова, О.П. Марковський // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – № 40. – 2020. – С.44-50. (Входить до наукометричних баз Open Academic Journals Index, Academic Resource Index ResearchBib, Rootindexing,

Information Matrix for the Analysis of Journals). *Автору належить метод відновлення інформації при втраті доступу до одного сховища, а також трьох блоків, які зберігаються на різних сховищах.*

2. Коляда К.В. Метод резервування та відновлення втрачених даних в глобальних мережах / К.В. Коляда, О.П. Марковський, В.Г. Саверченко, А.І. Торошанко // Телекомунікаційні та інформаційні технології. – № 1 (66) – 2020. – С.4-14. (Реферується наукометричними базами Google Scholar, Україніка наукова, а також входить до міжнародної SCOPUS наукометричної бази РІНЦ – Російського індексу наукового цитування). *Автору належить спосіб прискореного відновлення втрачених інформаційних блоків.*

3. Koliada K.V. Usage of linear erasure codes for increasing reliability and efficiency of information delivery on the internet / K.V. Koliada, N.G. Bardis, O.P. Markovskyi // International Journal of Circuits, Systems and Signal Processing – 2019 – Vol.13. – P.585-592. (Входить до наукометричних баз ISI Thomson's Scientific and Technical Proceedings, ISTP/ISI Proceedings, EI's Engineering Information Index, EI Compendex, DBLP, ACM, , Google Scholar та IEEE Xplore). *Автору належить вдосконалення методу відновлення затриманих при передачі інформаційних пакетів даних, за рахунок використання системи пріоритетів при виборі компонентів матриці формування резервних пакетів.*

4. Дичка И.А. Сравнительный анализ надежности запоминающих устройств с аппаратурной и кодовой избыточностью / И.А. Дичка, К.В. Коляда // Вестн. Киевского политехн. ин-та. Сер. Автоматика и электроприборостроение. – К., 1990. – № 27. – С. 35-38. (Входить до міжнародної наукометричної бази DOAJ –Directory of Open Access Journals, а також до бази РІНЦ – Російського індексу наукового цитування). *Автору належить обґрунтування вибору критеріїв ефективності резервування даних та порівняльний аналіз, з позиції обраних критеріїв, відновлення даних з використанням резервування та корегуючих кодів.*

5. Дичка И.А. Способ коррекции модульных ошибок в запоминающих устройствах / И.А. Дичка, К.В. Коляда, А.В. Наконечный // Вестн. Киевского политехн. ин-та. Сер. Автоматика и электроприборостроение. – К., 1990. – № 27. – С. 51-54. (Входить до міжнародної наукометричної бази DOAJ –Directory of Open Access Journals та до бази РІНЦ – Російського індексу наукового цитування). *Автору належить спосіб використання попередньо створених таблиць специфікацій для ефективного відновлення втраченої інформації.*

6. Дичка И.А. Сравнительный анализ надежности ЗУ с использованием различных корректирующих кодов / И.А. Дичка, Е.Ф. Колесник, К.В. Коляда // Вестн. Киевского политехн. ин-та. Сер. Автоматика и электроприборостроение. – К., 1989. – № 26. – С. 46-50. (Входить до міжнародної наукометричної бази DOAJ –Directory of Open Access Journals, а також до бази РІНЦ – Російського індексу наукового цитування). *Автору належить порівняльний аналіз відновлюючих та корегуючих кодів.*

7. Коляда К.В. Организация структур ПЗУ с сжатием данных / К.В.Коляда, И.А.Дичка // Вестн. Киевского политехн. ин-та. Сер. Автоматика и электроприборостроение. – К., 1988. – № 25. – С. 30-34. (Входить до міжнародної наукометричної бази DOAJ – Directory of Open Access Journals, а також до бази РІНЦ - Російського індексу наукового цитування). *Автору належить підхід до відновлення всіх блоків даних користувача, що зберігаються на одному сховищі.*

8. Коляда К.В. Применение обобщенного модульного кода Хемминга в компьютерных системах / К.В. Коляда, И.А. Дичка // Міжнародна науково-практична конференція "Проблеми інформатики та комп'ютерної техніки" (ПІКТ – 2013), Чернівецький національний університет імені Юрія Федьковича: Тези доповідей – Чернівці, 27-31 травня 2013 р. – С. 86-88. *Автору належить метод прискорення відновлення втраченої при віддаленому зберіганні інформації за рахунок розрідження матриці формування резервних блоків даних.*

9. Дичка И.А. Применение обобщенного кода Хемминга для обеспечения помехоустойчивости информационных систем / И.А. Дичка, К.В. Коляда, Е.С. Сулема // Тези доповідей 3-ї Української конференції з автоматичного керування "Автоматика-96". – Севастополь: СевГТУ, 1996. – Ч. II, С. 165-166. *Автору належить підхід до відновлення втрачених пачок пакетів даних в мережах.*

АНОТАЦІЇ

Коляда Костянтин Вячеславович. Методи та засоби підвищення ефективності відновлення даних, втрачених при їх віддаленому зберіганні та передачі в мережах. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.05 – Комп'ютерні системи та компоненти. – Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, 2021.

Дисертація присвячена проблемі підвищення ефективності відновлення даних, втрачених при їх віддаленому зберіганні в розподілених системах або при передачі глобальною мережею.

В роботі проведено аналіз факторів, що впливають на ефективність відновлення даних, втрачених під час їх довготривалого зберігання на віддалених розподілених системах, а також в процесі їх передачі глобальною мережею. Відновлення втрачених під час зберігання даних здійснюється за рахунок резервних даних, які також зберігаються на віддалених розподілених системах. Відновлення втрачених, або затриманих понад критичний проміжок часу даних при передачі в мережах також здійснюється шляхом передачі резервної інформації разом з основною.

Основний акцент в дисертаційному дослідженні зроблено на підвищенні швидкості відновлення втрачених даних та зменшенні об'єму резервних

даних за рахунок урахування особливостей реальних систем віддаленого зберігання даних користувачів.

Для прискорення відновлення втрачених при віддаленому зберіганні даних розроблено метод, оснований на використанні розріджених матриць формування резервних блоків. В реальних системах віддаленого зберігання інформації користувачів та системах комп'ютерного управління віддаленими об'єктами з використанням Інтернету в якості середовища обміну даних, важливу роль відіграє час відновлення втрачених інформаційних блоків чи затриманих доставкою понад критичний час пакетів даних.

Запропоновано ефективне технологічне рішення відновлення інформаційних блоків на основі попередньо створених таблиць специфікацій. За рахунок оптимізації вибору варіанту реконструювання і використання передобчислень досягається прискорення на 10-15% відновлення втрачених блоків.

Ключові слова: віддалене зберігання даних, хмарні технології, відновлюючі коди, лінійні коди, відновлення втрачених пакетів в мережах, реконструюючі коди, надійність передачі даних в глобальних мережах в реальному часі.

Koliada Kostiantyn Viacheslavovych. Method and tools for increasing the efficiency of recovering of data lost during remote storage and transmission in networks.- Manuscript.

Thesis for a PhD. degree by specialty 05.13.05 – Computer system and components. National Technical University of Ukraine "Igor Sykorsky Kiev Polytechnic Institute", Kyiv, 2021.

The dissertation is devoted to the problem of increasing the efficiency of data recovery, lost during their remote storage in distributed systems or during transmission over global networks.

The analysis of factors affecting for efficiency of data recovery lost during their long-term storage on remote distributed systems, as well as in the process of their transmission over global networks is carried out in the work. Recovery of data lost during storage is carried out at the expense of backup data, which is also stored on remote distributed systems. Recovery of lost or delayed over a critical period of time data during transmission in networks is also carried out by transmitting backup information along with the main.

The main emphasis in the dissertation research have been maid on increasing the speed of recovery of lost data and reducing the amount of backup data by taking into account the features of real systems for remote storage of user data.

To accelerate the recovery of data lost during remote storage, a method based on the use of sparse matrices for the formation of backup blocks has been developed. In real systems for remote storage of user information and computer management systems for remote objects using the Internet as a medium of data exchange, the time of recovery of lost information blocks or delayed delivery of critical data packets plays an important role.

An effective technological solution for the recovery of lost information blocks on the basis of previously created tables of specifications is proposed. Due to the

optimization of the choice of the reconstruction option and the use of precomputations, the acceleration of lost blocks recovery by 10-15% is achieved.

Keywords: distributed data storage, cloud technologies, erasured codes, linear codes, reconstruction lost packet in network, recoverable codes, reliability of data transmission in global network in real time.

Коляда Константин Вячеславович. Методы и средства повышения эффективности восстановления данных, утраченных при их удаленном хранении и передаче по сетям. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.05 – Компьютерные системы и компоненты. – Национальный технический университет Украины "Киевский политехнический институт имени Игоря Сикорского", Киев, 2021.

Диссертация посвящена проблеме повышения эффективности восстановления данных, потерянных при их отдаленном хранении в распределенных системах или при передаче по глобальным сетям.

В работе проведен анализ факторов, влияющие на эффективность восстановления данных, потерянных во время их длительного хранения на удаленных распределенных системах, а также в процессе их передачи по глобальным сетям. Восстановление утраченных во время хранения данных осуществляется за счет резервных данных, которые также хранятся на удаленных распределенных системах. Восстановление утраченных или задержанных более критического промежутка времени данных при передаче в сетях также осуществляется путем передачи резервной информации вместе с основной.

Основной акцент в диссертационном исследовании сделан на повышении скорости восстановления утраченных данных и уменьшении объема резервных данных за счет учета особенностей реальных систем удаленного хранения данных пользователей.

Для ускорения восстановления утраченных при удаленном хранении данных разработан метод, основанный на использовании разреженных матриц формирования резервных блоков. В реальных системах удаленного хранения информации пользователей и системах компьютерного управления удаленными объектами с использованием Интернета в качестве среды обмена данных, важную роль играет время восстановления потерянных информационных блоков или задержанных доставкой более критического времени пакетов данных.

Теоретически обосновано, разработан и исследован метод восстановления "пачек" пакетов и небольшого количества отдельных пакетов, который отличается способом построения матрицы формирования резервных пакетов, число которых за счет специализации на определенном доминирующем типе потерь пакетов, уменьшено по сравнению с известными методами.

Суть предлагаемого метода построения матрицы формирования резервных пакетов заключается в том, что они разделяются на две группы.

Единицы в строках первой группы находятся на расстоянии, равном максимальной длине пачки пакетов. Это позволяет разредить матрицу и, соответственно уменьшить количество операций, для восстановления потерянных пакетов.

Теоретическими и экспериментальными исследованиями установлено, что разработанный метод, за счет его ориентации на определенном доминирующем типе потерь информационных пакетов позволяет, в среднем на 30-40% сократить количество резервных пакетов, передаваемых по глобальным сетям. Предложенный в рамках метода способ построения матрицы формирования резервных пакетов обеспечивает, при этом сокращение времени восстановления потерянной "пачки" информационных пакетов в 2-4 раз, в зависимости от длины "пачки", за счет разрежения матрицы, что влечет уменьшение вычислительной сложности операций, связанных с их реконструированием.

Предложено эффективное технологическое решение восстановления утраченных информационных блоков на основе предварительно созданных таблиц спецификаций. За счет оптимизации выбора варианта реконструкции и использования передвычислений достигается ускорение на 10-15% восстановления утраченных блоков.

Ключевые слова: распределенное хранение данных, облачные технологии, восстанавливающие коды, линейные коды, восстанавливающие коды, восстановление утраченных пакетов в глобальных сетях, реконструирующие коды, надежность передачи данных в глобальных сетях в реальном времени.